

# An Approach for Similarity Matching and Comparison in Content based Image Retrieval System

**Er. Numa Bajaj**

CGC College of Engineering, Landran, CSE Department, Mohali, 140110, India  
E-mail: numabajaj@gmail.com

**Er. Jagbir Singh Gill and Dr. Rakesh Kumar**

CGC College of Engineering, Landran, CSE Department, Mohali, 140110, India and Sachdeva Engg. College for Girls, Gharuan, Mohali, India  
E-mail: cgccoe.cse.jagbir@gmail.com, rakesh77kumar@yahoo.com

**Abstract**—Today, in the age of images and digitization relevant retrieval is quite a topic of research. In past era, the database was having only text or database was low dimensional type. But with every new day thousands of pictures are getting added into the database making it a high dimensional data set. Therefore, from a high dimensional dataset to get a set of relevant images is quite a cumbersome task. Number of approaches for getting relevant retrieval is defined, some includes retrievals only on the basis of color, while some include more than one primitive feature to retrieve the relevant image such as color, shape and texture. In this paper experiment has been performed on the trademark images. Trademark is a very important asset for any organization and increasing trademark images have developed a quick need to organize these images. This paper includes the implementation of HSV model for fast retrieval. Which use color and texture so as to extract feature vector. Experiment takes query image and retrieve twelve most relevant images to the user. Further for performance evaluation parameter used is Precision and Recall.

**Index Terms**—CBIR, HSV, Recall and Precision, Searching.

## I. INTRODUCTION

With the advancement in technology, cameras have become inexpensive and sharing of information through images has increased, therefore, this has results into increase in number of images. Initially, database was used to be of low-dimensional data type, but with the increase in the number of images dramatically, it has now changed to high-dimensional type database. Therefore, keeping this in mind the old approach, which used to retrieve the results only by text, has now been replaced with new approach which is content based image retrieval (CBIR). It is a cumbersome job to make use of relevant data out of a huge database, which is the order of terabytes unless we have a system to organize it. Initially,

images were retrieved in CBIR system only on the basis of color feature of image. But with time and increased number of images using only color feature become an outdated technique and insufficient to meet the demands of modern era. After that various author introduced novel approaches so as to make this process of retrieval fast and efficient. Some of them uses combination of two or more primitive features i.e. shape, color and texture. Some uses special color model to enhance the performance. Therefore, CBIR system is a topic where already a lot of research has been done, but still attracts the researchers to make it more fast and accurate then it was ever before.

CBIR system adopts the following two step approach to search images from database [1].

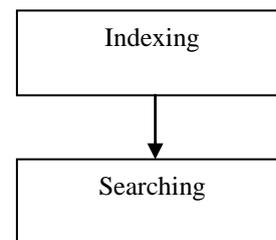


Fig. 1. Core Phases of CBIR system

- Indexing:** In this step, for every image in the database a feature vector is calculated capturing certain essential properties of the image. Further, feature vector is stored in the database along with the respective image.
- Searching:** In this step, given a query image whose feature vector is calculated, and it is compare with the feature vector of the images available in the database, and the most relevant results are retrieved and returned to user.

Some of the primary image contents are color and texture information [2],[3]

- Color:** Color is visual and most widely used

image content which is robust to all transformation and rotation [4],[3]. A user can easily differentiate between two images with the help of color feature. That is why, that is why this is most widely used feature in CBIR system. The most widely used method to represent color feature of an image is color histogram [5].

- b) Texture: This is a low level feature, mostly used to define the texture classification in an image.[5]

By extracting these primary contents the feature vector can be formed.[5]

In this paper, histogram based bins approach has been used for feature extraction. It combines texture and color contents to extract feature vector so as to increase its discriminating power. The presented approach uses HSV as color space model. For the next step, which is similarity matching between two images, Euclidean distance is used. This will results in relevant retrieval with respect to the query image. To evaluate the performance of our proposed approach which is HSV we have used Precision and Recall cross over point.

Rest of the paper is organized as follows; Section II contains brief explanation of HSV color space model. Section III explains the related work. Section IV explains the proposed work. Section V contains experimental test cases. Section VI contains the parameters for performance evaluation.

## II. HSV COLOR SPACE

HSV color space is one which is always preferred by any user for manipulating the color. This color space model is designed for better human perception.HSV color model is a cylindrical coordinate system. Colors are represented three basic features: hue, saturation and value [6].

**Hue:** The hue (H) represents the dominant spectral component—color in its pure form, as in green, red, or yellow.[7] Hue is measured by angle.



Fig. 2. Colors of Hue

**Saturation:** The Saturation (S) corresponds to adding white to the pure color changes the color: the less white, the more saturated the color is. [7] Saturation is measured by percentage.

**Value:** The value (V) corresponds to the brightness of

color. [7] It is also measured in percentage.

Hue, Saturation and Value can be represented using a inverted cone as shown below in the diagram. [8]

HSV color space is a non-linear but reversible conversion of RGB model. This conversion is important because HSV color space is more suitable for human perception in comparison to very basic RGB model. Following are the formulas using which RGB color space can be converted to HSV color space.[9]

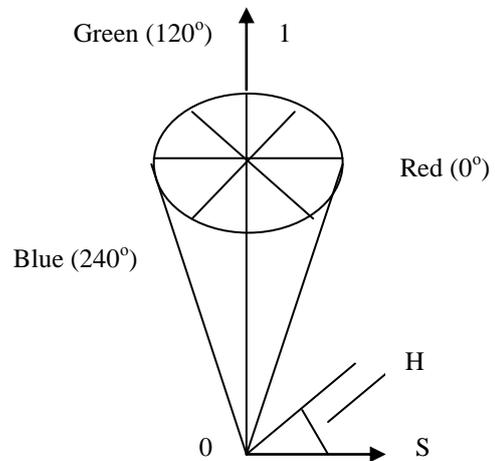


Fig. 3. HSV color space model

$$H = \arctan \frac{\sqrt{3}(G - B)}{(R - G) + (R - B)} \quad (1)$$

$$S = 1 - \frac{\min\{R, G, B\}}{V} \quad (2)$$

$$V = \frac{(R + G + B)}{3} \quad (3)$$

Figure 4 illustrates how RGB is converted to HSV graphically. [10] Suppose a point P is taken inside the triangle. All the three vertices of triangle show the three primary colors i.e. Red, Blue and Green.

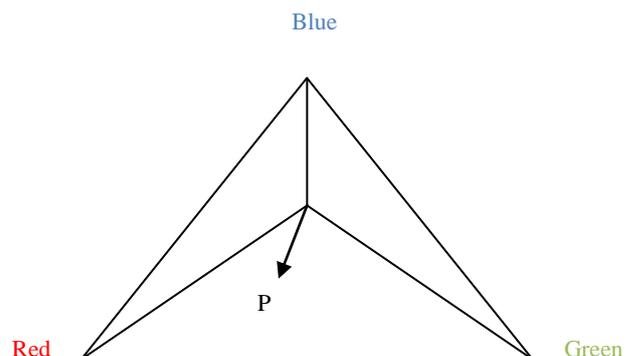


Fig. 4. RGB to HSV Conversion

**Hue:** Hue is defined by the angle between the line connecting point P to the centre and the line connecting

Red vertex to the centre.

**Saturation:** Saturation is defined as the distance between the Point P and the centre of triangle.

**Value:** It is also known as intensity. It is defined as the height of the line perpendicular the triangle passing through its centre.

### III. RELATED WORK

Content based image retrieval is a topic on which research is being carried out from last many years. With content based image retrieval there was a problem in similarity matching because the color histograms were using only frequency value of the same color after performing color quantization, which results in quantization error.

This problem was solved by Y. J., Park, where to reduce the error of color quantization mean value of RGB color component and color frequency for each region was calculated separately, so as to retrieve the most similar images.[3]

All the image retrievals were only on the basis o color. Hiremath, gives a novel approach in which for image retrieval primitive image descriptors were used i.e. color, texture and shape, so as to achieve better results of retrieval. For this author, uses color moments, which serves as local descriptor to the features, color and texture. Shape feature is calculated using gradient vector flow fields. Combination of these three features results into better performance than before.[2]

Murala, represented a novel approach of image retrieval using only color and texture features. In this approach the retrieval results were calculated using color histogram and Gobar wavelet transform.[4]

Arthi, proposed a new approach for efficient image retrieval, which uses a algorithm based on CCM (Color Co-occurrence matrix). For calculating CCM, HSV model is used.[10]

Kekre, introduced a novel idea of extracting the feature vectors of image along with dimensionality reduction. The author implemented the bins approach which is designed and implemented with the help of image histogram. This bins extraction results in reducing the feature vector dimension and efficiency of retrievals are improved.[11]

Kekre, enhanced the approach of bins extraction. The author execute the bins approach in various color spaces which results in positive change, which was not seen with RGB before. The work was performed using two primitive features i.e. color and texture associated with 8-bins approach which results in dimensionality reduction.

### IV. PROPOSED WORK

Two major problems which attracts researchers to explore more in CBIR system is:

- Huge database which needs to organized properly so that quick and accurate results can be retrieved.

- Information System suffers from Sensory Gap and Semantic Gap due to difference in machine algorithms.

Therefore, these two problems lead to the following objective:

- To develop a CBIR system for better and accurate results using HSV model (Best model for user perception).
- a) Basic Design for the Proposed Work

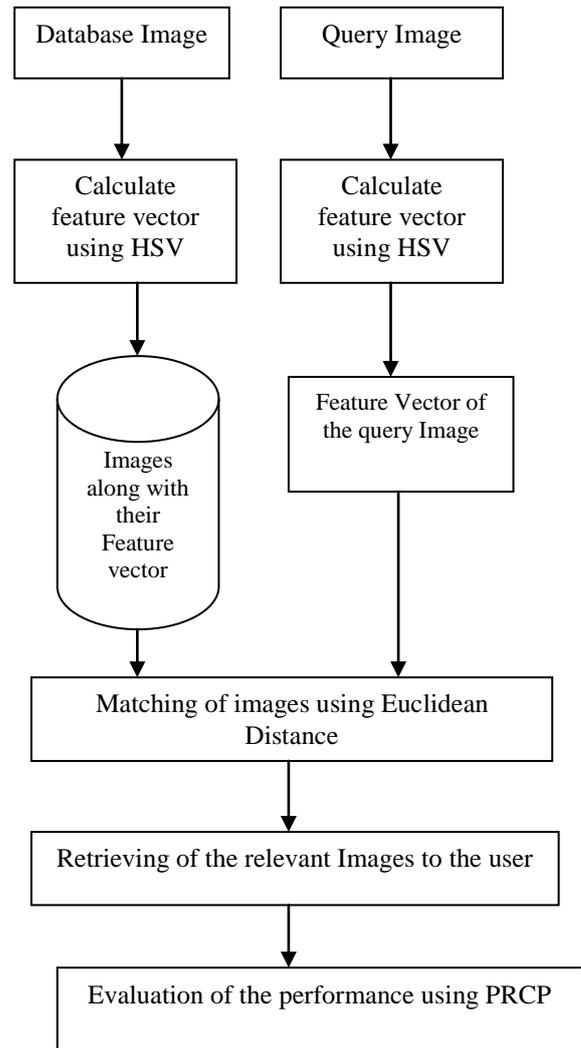


Fig. 5. Basic Design of the CBIR approach

**Database Images:** These are the images present in database from which relevant images will be retrieved. For this proposed work we have taken set of 50 images, which are trademark images.

**Query Image:** This is the image which user send to the machine to get relevant results from the images present in the database.

**Feature Extraction:** Feature is a very important part of any image. It uniquely identifies an image. Therefore for every image we have to extract its feature vector so as to perform similarity matching between indexed images

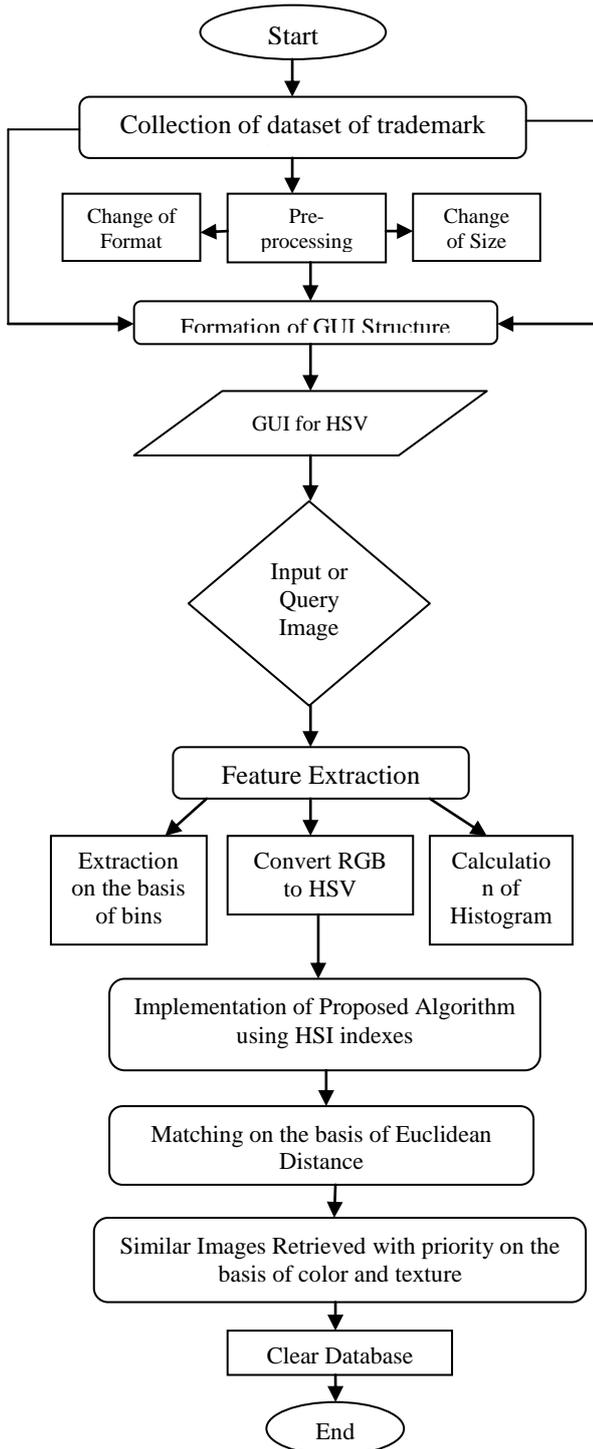


Fig. 6. Flow Chart for the Proposed Work

and the query image. For this proposed work we are mainly using color and the texture features for preparing feature vectors.

**Similarity Matching:** Once features have been extracted from the images, after that the query image is compared with the images present in the database. For the proposed work we are using Euclidean distance to find the similarity between two images. Formula for Euclidean distance is:[10]

$$Euclidean \ Distance = \sqrt{\sum_{i=1}^n [Q_i - D_i]^2} \quad (4)$$

**Relevant Images:** These are the images retrieved which are most relevant to the query image. For proposed work we have taken 12 slots so as to retrieve relevant images. But this is not necessary that all the images are relevant, some of the non relevant images can also be retrieved.

**Performance Evaluation:** Performance will be evaluated using PRCP for this paper.

b) Algorithm Level Design for Proposed Work

V. EXPERIMENTAL TEST CASES

Distance between two images is computed as where, the Euclidean distance of features of the two images is. The database used by us consists of 50 groundtruthed images. It describes each individual image by a set of phrases or keywords. Performance of the system is studied by using each image in the database as the query image and top 12 similar images are retrieved by Euclidean distance based exhaustive search.

Query Image: Is the image which is fired to the system.

For this system, this google trademark image has been fired, so as to get results relevant to it.



Fig. 7. Query Image

Retrieved image: These are the images which system will retrieve on the basis of Euclidean distance. Here in the diagram below 12 retrieved images are shown with respect to the query image fired to the system. These 12 images are most relevant images out of the total 50 images in database. This retrieval is based on the color and texture of the query image. Relevance of the image decreases as we move from left to right.

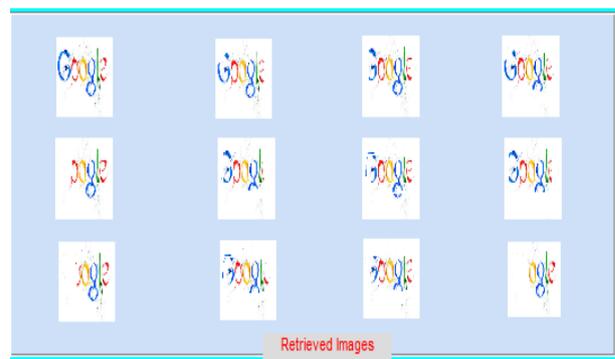


Fig. 8. Retrieved Images

Various Test Cases are:-



Fig. 9. Query Image



Fig. 10. Retrieved Images

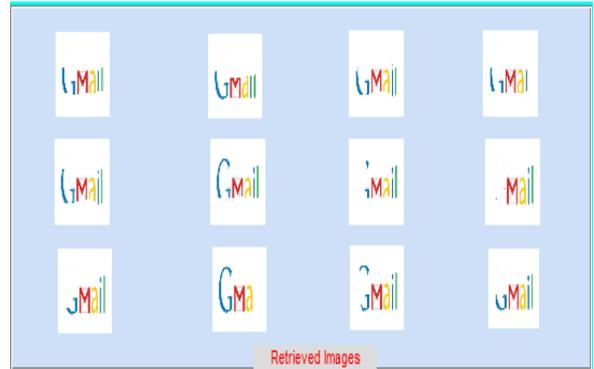


Fig. 14. Retrieved Images



Fig. 15. Logo Images



Fig. 11. Query Image

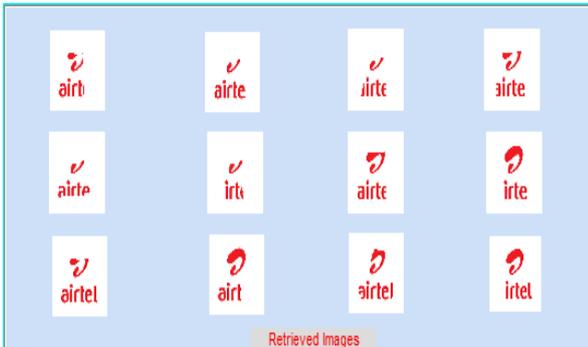


Fig. 12. Retrieved Images



Fig. 13. Logo Images

### VI. PERFORMANCE EVALUATION

Finally, for evaluating the performance of HSV model Precision and Recall graph are as follows.

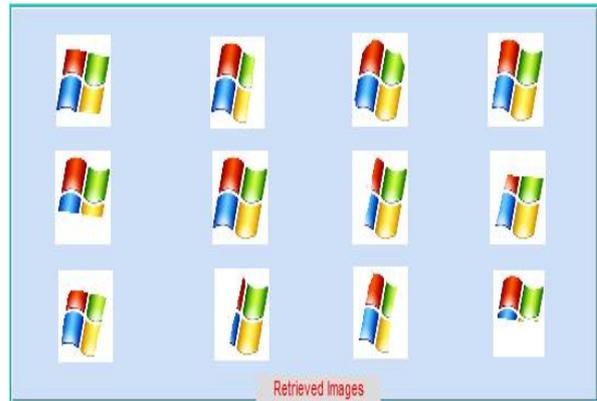


Fig. 16. Retrieved Images

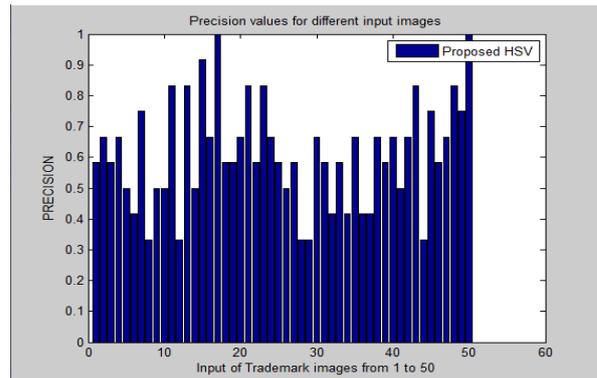


Fig. 17. Precision Graph for HSV model

Fig 17 graph shows Precision on Y-axis and Input of Trademark images 1 to 50 on X-axis. It clearly depicts the proposed HSV (HSI) method on the basis of precision value. For the proposed algorithm the precision value is quiet better and optimist which can be seen here from the above figure. Higher will be the precision value higher will be the accuracy of results. The precision value calculated on the basis on number of favourable retrieved image with respect to all retrieved images.

Fig 18 shows Recall on Y-axis and Input of Trademark images on X-axis. It clearly depicts the proposed HSV (HSI) method on the basis of recall value. For the proposed algorithm the recall value is quiet better and optimist which can be seen here from the above figure. Higher recall value shows that the results are accurate. The recall value is calculated on the basis on number of favorable retrieved image with respect to all associated and similar images in the datasets.

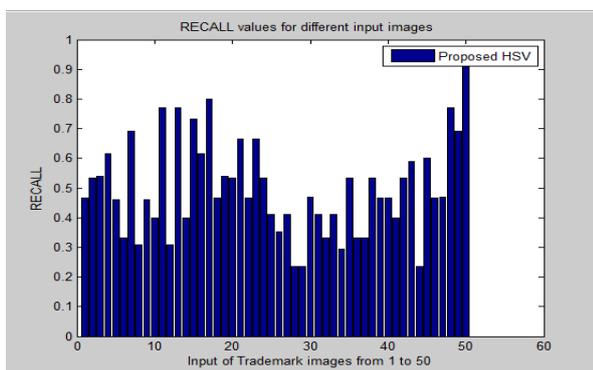


Fig. 18. Recall Graph for HSV Model

## VII. CONCLUSION AND FUTURE SCOPE

In this paper, stress has been laid towards the CBIR of Trademark Images using HSV model. From the various primary properties of an image, we have chosen color and texture for extracting the feature vector. The performance evaluation was made over each images stored in the set of data set of various Trademark images. Here, the parameter taken for performance evaluation is PRECISION and RECALL. The results clearly depicts that the value of PRECISION for all the images stored in the set of dataset is quiet higher for proposed HSV (HSI) approach. The result of proposed work is quiet full fill the above line.

Furthermore, there is need to improve the results of image retrieval more to the present one. There is also need to increase the work more for different techniques and also there should be need to use some more parameters for performance evaluation.

## REFERENCES

- [1] Huang, J., Kumar, S. R., Mitra, M., Zhu, W. J., & Zabih, R. (1999). Spatial color indexing and applications. *International Journal of Computer Vision*, 35(3), 245-268.
- [2] Hiremath, P. S., & Pujari, J. (2007, December). Content based image retrieval using color, texture and shape features. In *Advanced Computing and Communications, 2007. ADCOM 2007. International Conference on* (pp. 780-784). IEEE.
- [3] Song, Y. J., Park, W. B., Kim, D. W., & Ahn, J. H. (2004, November). Content-based image retrieval using new color histogram. In *Intelligent Signal Processing and Communication Systems, 2004. ISPACS 2004. Proceedings of 2004 International Symposium on* (pp. 609-611). IEEE.
- [4] Murala, S., Gonde, A. B., & Maheshwari, R. P. (2009, March). Color and texture features for image indexing and retrieval. In *Advance Computing Conference, 2009. IACC 2009. IEEE International* (pp. 1411-1416). IEEE.
- [5] Kekre, H. B., & Sonawane, K. (2014, April). Comparative study of color histogram based bins approach in RGB, XYZ, Kekre's LXY and L' X' Y' color spaces. In *Circuits, Systems, Communication and Information Technology Applications (CSCITA), 2014 International Conference on* (pp. 364-369). IEEE.
- [6] Ketenci, S., & Gencturk, B. (2013, July). Performance analysis in common color spaces of 2D Gaussian Color Model for skin segmentation. In *EUROCON, 2013 IEEE* (pp. 1653-1657). IEEE.
- [7] Manjunath, B. S., Ohm, J. R., Vasudevan, V. V., & Yamada, A. (2001). Color and texture descriptors. *Circuits and Systems for Video Technology, IEEE Transactions on*, 11(6), 703-715.
- [8] Zhao, Q., Yang, J., Yang, J., & Liu, H. (2009, April). Stone images retrieval based on color histogram. In *Image Analysis and Signal Processing, 2009. IASP 2009. International Conference on* (pp. 157-161). IEEE.
- [9] Yu, H., Li, M., Zhang, H. J., & Feng, J. (2002, June). Color texture moments for content-based image retrieval. In *Image Processing. 2002. Proceedings. 2002 International Conference on* (Vol. 3, pp. 929-932). IEEE.
- [10] Arthi, K., & Vijayaraghavan, M. J. (2013). Content based image retrieval algorithm using colour models. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(3), 1343-47.
- [11] Kekre, H. B., & Sonawane, K. (2012). Histogram Partitioning for Feature Vector Dimension Reduction in Bins Approach for CBIR. *IJECCCE*, 3(6), 1630-1639.
- [12] Kekre, H. B., & Sonawane, K. (2013). Performance evaluation of bins approach in YCbCr color space with and without scaling. *International Journal of Soft Computing and Engineering*, 3(3), 203-210.
- [13] Müller, W., Squire, D. M., Marchand-Maillet, S., & Pun, T. (2001). Performance evaluation in content-based image retrieval: overview and proposals. *Pattern Recognition Letters*, 22(5), 593-601.
- [14] Sharma, N. S., Rawat, P. S., & Singh, J. S. (2011). Efficient CBIR using color histogram processing. *Signal & Image Processing*, 2(1).
- [15] Suhasini, P. S., Krishna, K., & Krishna, I. M. (2009). CBIR USING COLOR HISTOGRAM PROCESSING. *Journal of Theoretical & Applied Information Technology*, 6(1).
- [16] Jeong, S., Won, C. S., & Gray, R. M. (2004). Image retrieval using color histograms generated by Gauss mixture vector quantization. *Computer Vision and Image Understanding*, 94(1), 44-66.
- [17] Datta, R., Li, J., & Wang, J. Z. (2005, November). Content-based image retrieval: approaches and trends of the new age. In *Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval* (pp. 253-262). ACM.
- [18] Schettini, R., Ciocca, G., & Zuffi, S. (2001). A survey of

methods for colour image indexing and retrieval in image databases. *Color Imaging Science: Exploiting Digital Media*, 183-211.

### Authors' Profiles



**Er. Numa Bajaj** is from Chandigarh. Born on 25 January 1992. She completed B.tech (Information Technology) from Rayat-Bahra Institute of Engineering and Bio-Technology, Mohali, India in the year 2013. She is pursuing M.tech (Computer Science) from CGC College of Engineering, Landran,

Mohali, India.



**Er. Jagbir Singh Gill** is from Chandigarh. Born on 1 April 1986. He completed his BE(Computer Science) from SLIET Longowal, India in the year 2010 and M.tech(Computer Science) from Gurunank Dev Engineering College, Ludhiana, India in the year 2013. He is pursuing his Phd.

(Computer Science) in Research area Medical Informatics.

Er. Gill has teaching experience of five years.

**How to cite this paper:** Er. Numa Bajaj, Er. Jagbir Singh Gill, Rakesh Kumar, "An Approach for Similarity Matching and Comparison in Content based Image Retrieval System", *IJIEEB*, vol.7, no.5, pp.48-54, 2015. DOI: 10.5815/ijieeb.2015.05.07